

IN '64
003507
541

NASA Contractor Report 191583
ICASE Report No. 93-95

ICASE



ON THE DAUBECHIES-BASED WAVELET DIFFERENTIATION MATRIX

Leland Jameson (NASA-CR-191583) ON THE
DAUBECHIES-BASED WAVELET
DIFFERENTIATION MATRIX Final Report
(ICASE) 54 p

N94-23090

Unclas

G3/64 0203567

NASA Contract No. NAS1-19480
December 1993

Institute for Computer Applications in Science and Engineering
NASA Langley Research Center
Hampton, Virginia 23681-0001

Operated by the Universities Space Research Association



National Aeronautics and
Space Administration
Langley Research Center
Hampton, Virginia 23681-0001

ON THE DAUBECHIES-BASED WAVELET DIFFERENTIATION MATRIX

*Leland Jameson*¹

Institute for Computer Applications in Science and Engineering
NASA Langley Research Center
Hampton, VA 23681

ABSTRACT

The differentiation matrix for a Daubechies-based wavelet basis will be constructed and ‘superconvergence’ will be proven. That is, it will be proven that under the assumption of periodic boundary conditions that the differentiation matrix is accurate of order $2M$, even though the approximation subspace can represent exactly only polynomials up to degree $M - 1$, where M is the number of vanishing moments of the associated wavelet. It will be illustrated that Daubechies-based wavelet methods are equivalent to finite difference methods with grid refinement in regions of the domain where small-scale structure is present.

¹This research was supported by the National Aeronautics and Space Administration under NASA Contract No. NAS1-19480 while the author was in residence at the Institute for Computer Applications in Science and Engineering (ICASE), NASA Langley Research Center, Hampton, VA 23681. Research was also supported by AFOSR grant 93-1-0090, by DARPA grant N00014-91-4016, and by NSF grant DMS-9211820.

1 Introduction

The term *differentiation matrix* was coined by E. Tadmor in his review on spectral methods [1]. The term denotes the transformation between grid point values of a function and its approximate first derivative. This matrix is a product of three matrices.

The first matrix C is constructed as follows: assume that the point values of a function $f(x)$ (where $a \leq x \leq b$) are given at N points x_j for $0 \leq j \leq N - 1$. Thus a vector of numbers $f(x_j)$ is given. From this vector one can reconstruct an approximation to the function $f(x)$ for every point x in the interval. This approximation (denoted by $P_N f$) itself belongs to a finite dimensional space - in pseudospectral methods it is the global interpolation polynomial that collocates $f(x_j)$ and in finite differences or finite elements it is a piecewise polynomial. This transformation between $f(x_j)$ and $P_N f$, defines the matrix C . Of course this matrix depends on the special basis chosen to represent $P_N f$. A good example is the Fourier interpolation procedure in which the basis is the set of complex exponentials.

The second matrix D results from differentiating $P_N f$, and projecting it back to the finite dimensional space. Thus D is defined by the *linear* transformation between $P_N f$ and $P_N \frac{d}{dx} P_N f$.

The last matrix is the inverse of the first matrix. Basically, since we are given the approximation $P_N \frac{d}{dx} P_N f$ we can read it at the grid points x_j . Thus the differentiation matrix \mathcal{D} can be represented as $\mathcal{D} = C^{-1} D C$.

In this paper the wavelet differentiation matrix will be examined. As with other basis sets, as outlined above, it is a product of three matrices. Under the assumption of periodicity of $f(x)$, however, the matrices C and D commute allowing D to operate directly on the vector of numbers $f(x_j)$. That is, the differentiation matrix \mathcal{D} is simply D : $\mathcal{D} = D$. Furthermore, the matrix D differentiates samples of polynomials exactly, i.e., the action of D is equivalent to a finite difference operator with order of accuracy

depending on the order of the wavelet chosen.

More precisely, the following outlines the proof of this assertion:

- Given a periodic function $f(x)$, let C be the mapping from evenly-spaced samples of $f(x)$ to the approximate scaling function coefficients on the finest scale: $C : \vec{f} \rightarrow \vec{\sigma}_0$. Due to the periodicity of $f(x)$, C is circulant in form.
- Let D be the mapping from the exact scaling function coefficients of $f(x)$ to scaling function coefficients of $f'(x)$: $D : \vec{s}_0 \rightarrow \vec{s}'_0$. Once again, due to the periodicity of $f(x)$, D is circulant in form.
- The matrix operator D can differentiate exactly evenly-spaced samples of polynomials, i.e., D has the effect of a finite-difference operator. The order of exact differentiation depends on the order of the wavelet used.
- All circulant matrices with the same dimensions commute, therefore the operator D can be applied directly to \vec{f} :

$$\vec{f}' = C^{-1}DC\vec{f},$$

or simply,

$$\vec{f}' = D\vec{f},$$

and this will complete the proof.

This paper contains five sections:

§1) This introduction.

§2) Standard definitions of wavelets and scaling functions are given.

§3) The general approximation properties of wavelets will be discussed along with the quadrature formula needed to approximate the scaling function coefficients of a function.

§4) This is the most important section of this paper. It will be proved that the action of D is equivalent to a finite difference operator.

§5) The results of sections (3) and (4) are combined for the desired conclusion.

In addition, the following two related topics are explored in the first two appendices:

Appendix A) For wavelets supported on $(0,3M)$ it will be shown that $\int \phi(x)x^m dx = (\int \phi(x)xdx)^m$.

Appendix B) The moments of the scaling function $\phi(x)$ will be calculated.

2 Wavelet Definitions and Relations

The term wavelet is used to describe a spatially localized function. ‘Localized’ means that the wavelet has compact support or that the wavelet almost has compact support in the sense that outside of some interval the amplitude of the wavelet decays exponentially. We will consider only wavelets that have compact support and that are of the type defined by Daubechies [2] which are supported on $[0, 2M - 1]$, where M is the number of vanishing moments defined later in this section.

To define Daubechies wavelets, consider the two functions $\phi(x)$ and $\psi(x)$ which are solutions to the following equations:

$$\phi(x) = \sqrt{2} \sum_{k=0}^{L-1} h_k \phi(2x - k), \quad (1)$$

$$\psi(x) = \sqrt{2} \sum_{k=0}^{L-1} g_k \phi(2x - k), \quad (2)$$

where $\phi(x)$ is normalized,

$$\int_{-\infty}^{\infty} \phi(x) dx = 1. \quad (3)$$

Let,

$$\phi_k^j(x) = 2^{-\frac{j}{2}} \phi(2^{-j}x - k), \quad (4)$$

and

$$\psi_k^j(x) = 2^{-\frac{j}{2}} \psi(2^{-j}x - k), \quad (5)$$

where $j, k \in Z$, denote the dilations and translations of the scaling function and the wavelet.

The coefficients $H = \{h_k\}_{k=0}^{L-1}$ and $G = \{g_k\}_{k=0}^{L-1}$ are related by $g_k = (-1)^k h_{L-k}$ for $k = 0, \dots, L - 1$. Furthermore, H and G are chosen so that dilations and translations of the wavelet, $\psi_k^j(x)$, form an orthonormal basis of $L^2(\mathcal{R})$ and so that $\psi(x)$ has M vanishing moments. In other words, $\psi_k^j(x)$ will satisfy

$$\delta_{kl} \delta_{jm} = \int_{-\infty}^{\infty} \psi_k^j(x) \psi_l^m(x) dx, \quad (6)$$

where δ_{kl} is the Kronecker delta function. Also, $\psi(x) = \psi_0^0(x)$ satisfies

$$\int_{-\infty}^{\infty} \psi(x)x^m dx = 0, \quad (7)$$

for $m = 0, \dots, M - 1$. Under the conditions of the previous two equations, for any function $f(x) \in L^2(\mathbb{R})$ there exists a set $\{d_{jk}\}$ such that

$$f(x) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} d_{jk} \psi_k^j(x), \quad (8)$$

where

$$d_{jk} = \int_{-\infty}^{\infty} f(x) \psi_k^j(x) dx. \quad (9)$$

The number of vanishing moments of the wavelet $\psi(x)$ defines the accuracy of approximation. The two sets of coefficients H and G are known in signal processing literature as quadrature mirror filters [3]. For Daubechies wavelets the number of coefficients in H and G , or the length of the filters H and G , denoted by L , is related to the number of vanishing moments M by $2M = L$. For example, the famous Haar wavelet is found by defining H as $h_0 = h_1 = 1$. For this filter, H , the solution to the dilation equation (1), $\phi(x)$, is the box function: $\phi(x) = 1$ for $x \in [0, 1]$ and $\phi(x) = 0$ otherwise. The Haar function is very useful as a learning tool, but it is not very useful as a basis function on which to expand another function for the important reason that it is not differentiable. The coefficients, H , needed to define compactly supported wavelets with a higher degree of regularity can be found in [2]. As is expected, the regularity increases with the support of the wavelet. The usual notation to denote a Daubechies wavelet defined by coefficients H of length L is D_L .

It is usual to let the spaces spanned by $\phi_k^j(x)$ and $\psi_k^j(x)$ over the parameter k , with j fixed, to be denoted by V_j and W_j respectively:

$$V_j = \mathop{\text{span}}_{k \in \mathbb{Z}} \phi_k^j(x), \quad (10)$$

$$W_j = \mathop{\text{span}}_{k \in \mathbb{Z}} \psi_k^j(x). \quad (11)$$

The spaces V_j and W_j are related by [2]

$$\dots \subset V_1 \subset V_0 \subset V_{-1} \subset \dots, \quad (12)$$

and that

$$V_j = V_{j+1} \oplus W_{j+1}. \quad (13)$$

The previously stated condition that the wavelets form an orthonormal basis of $L^2(\mathbb{R})$ can now be written as,

$$L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j. \quad (14)$$

Two final properties of the spaces V_j are that

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\}, \quad (15)$$

and

$$\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R}). \quad (16)$$

Properties of the Semi-Discrete Fourier Transform (SDFT) of the filter H will also be needed. The following definition is not exactly the SDFT but a constant times the SDFT:

$$\hat{H}(\xi) = 2^{-1/2} \sum_{k=0}^{L-1} h_k e^{ik\xi}. \quad (17)$$

This DFT satisfies the following equation, see [4]:

$$|\hat{H}(\xi)|^2 + |\hat{H}(\xi + \pi)|^2 = 1. \quad (18)$$

Solutions of equation (18) have the following properties, see [2]:

$$\hat{H}(\xi) = \left(\frac{1}{2}(1 + e^{i\xi})\right)^M Q(e^{i\xi}), \quad (19)$$

where M is the number of vanishing moments of the wavelet and Q is a trigonometric polynomial such that,

$$|Q(e^{i\xi})|^2 = P(\sin^2(\xi/2)) + \sin^{2M}(\xi/2)R\left(\frac{1}{2}\cos\xi\right), \quad (20)$$

where

$$P(y) = \sum_{k=0}^{M-1} \binom{M-1+k}{k} y^k, \quad (21)$$

and R is an odd polynomial such that,

$$0 \leq P(y) + y^M R(1/2 - y) \quad (22)$$

for $0 \leq y \leq 1$, and

$$\sup_{0 \leq y \leq 1} (P(y) + y^M R(1/2 - y)) < 2^{2(M-1)}, \quad (23)$$

if $M \geq 2$ or

$$-\frac{2}{1-2|x|} \leq R(x) \leq \frac{2}{1+2|x|}, \quad (24)$$

for $|x| \leq 1/2$, if $M = 1$. The important point here is that $\hat{H}(\xi)$ has a zero of order M at $\xi = \pi$.

Of course, infinite sums and unions are meaningless when one begins to implement a wavelet expansion on a computer. In some way one must limit range of the scale parameter j and the location parameter k . Consider first the scale parameter j . As stated above, the wavelet expansion is complete: $L^2(R) = \bigoplus_{j \in Z} W_j$. Therefore, any $f(x) \in L^2(R)$ can be written as,

$$f(x) = \sum_{j \in Z} \sum_{k \in Z} d_k^j \psi_k^j(x),$$

where due to orthonormality of the wavelets $d_k^j = \int_{-\infty}^{\infty} f(x) \psi_k^j(x)$. In this expansion, functions with arbitrarily small-scale structures can be represented. In practice, however, there is a limit to how small the smallest structure can be. This would depend, for example, on how fine the grid is in a numerical computation scenario or perhaps what the sampling frequency is in a signal processing scenario. Therefore, on a computer an expansion would take place in a space such as

$$V_0 = W_1 \oplus W_2 \oplus \dots \oplus W_J \oplus V_J, \quad (25)$$

and would appear as,

$$P_{V_0} f(x) = \sum_{k \in \mathbb{Z}} s_k^J \phi_k^J(x) + \sum_{j=1}^J \sum_{k \in \mathbb{Z}} d_k^j \psi_k^j(x), \quad (26)$$

where again due to orthonormality of the basis functions $d_k^j = \int_{-\infty}^{\infty} f(x) \psi_k^j(x)$, and $s_k^J = \int_{-\infty}^{\infty} f(x) \phi_k^J(x)$. In this expansion, scale $j = 0$ is arbitrarily chosen as the finest scale that is needed, and scale J would be the scale at which a kind of local average, $\phi_k^J(x)$, provides sufficient large scale information. In language that is likely to appeal to the electrical engineer it can be said that $\phi_k^J(x)$ represents the direct current portion of a signal and that $\psi_k^j(x)$ represents the alternating current portion of a signal at, very roughly, frequency j . As stated above, one must also limit the range of the location parameter k . In this paper this is done by assuming that $f(x)$ is a periodic function. The periodicity of $f(x)$ induces periodicity on all wavelet coefficients, s_k^j and d_k^j .

This completes the definition of wavelets. The next section will discuss function approximation in a wavelet basis.

3 Approximating in a Wavelet Basis

Scaling functions and wavelets were defined in the previous section. The goal of this section is to find the coefficients in a wavelet expansion. More precisely, the scaling function coefficients at the finest scale, \bar{s}_0 , will be approximated. The key to this approximation is the matrix C which maps evenly-spaced samples of a periodic function to the approximate scaling function coefficients. This matrix C has the desirable property of being circulant in form with the ramification that C will commute with any other circulant matrix, particularly the derivative matrix R^0 , the subject of section (4). An example of C is given at the end of section (3.2).

The scenario for this section is as follows: let the finest scale be scale $j = 0$, i.e., at this scale all relevant small scale structures in the function have been captured and represented. One seeks an expansion of a function $f(x)$ in terms of ϕ_k^0 in the space V_0 . With the projection P_{V_0} defined as $P_{V_0} : L^2(\mathbb{R}) \rightarrow V_0$ such an expansion has the following form:

$$P_{V_0}f(x) = \sum_{k \in \mathbb{Z}} s_k^0 \phi_k^0(x), \quad (27)$$

where due to the orthonormality of the basis functions, $\delta_{ij} = \int_{-\infty}^{\infty} \phi_i^0(x) \phi_j^0(x) dx$, the coefficients s_k^0 are given by

$$s_k^0 = \int_{-\infty}^{\infty} f(x) \phi_k^0(x) dx. \quad (28)$$

Once the s_k^0 have been found one would usually then find the scaling function and wavelet function coefficients at more coarse scales. This can be done by using equation (29) to get s_k^j for $j = 1, \dots, J$ and by using equation (30) to get d_k^j for $j = 1, \dots, J$. These equations are derived respectively from equations (1) and (2), see [4], [5],

$$s_k^j = \sum_{n=1}^{2M} h_n s_{n+2k-2}^{j-1}, \quad (29)$$

and

$$d_k^j = \sum_{n=1}^{2M} g_n s_{n+2k-2}^{j-1}. \quad (30)$$

In this section, however, the decomposition onto more coarse scales will not be calculated. The important step for this section is the approximation of the integral $s_k^0 = \int_{-\infty}^{\infty} f(x)\phi_k^0(x)dx$. Let σ_k^0 denote the approximation to s_k^0 . The quadrature formula for this integral encompasses the approximation properties of scaling functions, and hence wavelets.

This section contains 3 subsections:

§3.1) The approximation properties of scaling functions will be discussed and the quadrature formula to estimate the integral $\int_{-\infty}^{\infty} f(x)\phi_k^0(x)dx$ will be derived.

§3.2) An example using the results from section (3.1) is given for the Daubechies wavelet D_6 .

§3.3) The example from section (3.2) leads to a circulant matrix for the matrix C . Circulant matrices will be defined and the ramifications of circularity will be discussed.

3.1 Quadrature Formula for Scaling Function

In this subsection the coefficients s_k^0 will be approximated. Before stating the appropriate quadrature formula, however, the order of accuracy of a wavelet approximation is discussed.

3.1.1 Approximation Properties of Scaling Functions

This subsection comes essentially from [6]. The approximation properties of scaling functions are determined by the Discrete Fourier Transform of the filter H . That is, if

$$\hat{H}(\xi) = \frac{1}{2} \sum_{k=0}^{L-1} h_k e^{ik\xi} \quad (31)$$

has a zero of order M at $\xi = \pi$ then there are a number of consequences:

1. The polynomials $1, x, \dots, x^{M-1}$ are linear combinations of the translates of the scaling function ϕ_k^0 .

2. Smooth functions can be approximated with error $O(h^M)$, where h denotes the grid size, i.e., there exist a set s_k^j , where j is fixed, and there exists a constant C such that

$$|f(x) - \sum_k s_k^j \phi_k^j(x)| \leq Ch^M |f^{(M)}(x)|, \quad (32)$$

where the norm $|\cdot|$ is the L_2 norm.

3. The associated wavelet has M vanishing moments,

$$\int x^m \psi(x) dx = 0$$

for $m = 0, \dots, M - 1$.

Other ramifications can be found in [6]. These approximation properties determine the accuracy of the quadrature formula used to approximate the scaling function coefficients \vec{s}_0 which is derived in the following section.

3.1.2 Derivation of Quadrature Formula

It is important to note that all wavelets in this paper are of the usual Daubechies type, i.e., the support of a usual Daubechies wavelet D_{2M} is $[0, 2M - 1]$ where M is the number of vanishing moments of the wavelet. For this subsection this support size is particularly important to keep in mind because there does exist an orthonormal family of wavelets which are supported on $[0, 3M - 1]$ and which have a very simple quadrature formula based on the vanishing moments of the wavelet (see appendix A) but this is not the wavelet being used in this paper.

Given the approximation properties of the scaling function from the previous subsection, one can now seek a quadrature formula which is exact when $f(x)$ is a polynomial up to order $M - 1$: $f(x) = p(x) \in \mathbf{P}_{M-1}$. That is, there exist a set of coefficients $\{c_l\}_{l=0}^{M-1}$ such that

$$\int_{-\infty}^{\infty} p(x) \phi_k^0 dx = \sum_{l=0}^{M-1} c_l p(l+k), \quad (33)$$

for $p(x) \in \mathbf{P}_{M-1}$. If the integral is shifted the above equation becomes,

$$\int_{-\infty}^{\infty} p(y+k)\phi_0^0(y)dy = \sum_{l=0}^{M-1} c_l p(l+k). \quad (34)$$

More simply, the coefficients $\{c_l\}_{l=0}^{M-1}$ can be found [9] by solving the following linear system:

$$\int_{-\infty}^{\infty} x^m \phi(x) dx = \sum_{l=0}^{M-1} l^m c_l, \quad (35)$$

for $m = 0, 1, \dots, M-1$, and the coefficients $\{c_l\}_{l=0}^{M-1}$ provide the desired quadrature formula. That is, the coefficients, $\vec{\sigma}_0$, which approximate \vec{s}_0 are found from,

$$\sigma_k^0 = \sum_{l=0}^{M-1} c_l f(l+k). \quad (36)$$

When placed in matrix form the coefficients $\{c_l\}_{l=0}^{M-1}$ yield the circulant matrix C . A more thorough discussion of circulant matrices will be given in §(3.3) after the example of the next subsection has been completed.

Note that since the above derived quadrature formula is exact for $p(x) \in \mathbf{P}_{M-1}$ the coefficients σ_k^0 approximate the coefficients s_k^0 with error of order M . Also, note that the derivation of the quadrature coefficients depends only on the moments of the scaling function, $\int_{-\infty}^{\infty} x^m \phi(x) dx$. In the next subsection, the moments of the scaling function will first be calculated and then the coefficients $\{c_l\}_{l=0}^{M-1}$ will be found for the D_6 wavelet. The wavelet D_6 is chosen for no other reason than that D_2 and D_4 receive considerable attention from other sources and that D_6 is slightly less trivial than D_2 and D_4 while remaining manageable.

3.2 Example with the Daubechies Wavelet D_6

Recall from the previous subsection that the immediate goal is to approximate the scaling function coefficients of a function at scale $j = 0$. Specifically, in this section the objective is to derive the matrix form of the mapping from evenly-spaced samples of a periodic function $f(x)$ to the scaling function coefficients on the finest scale s_k^0 . The example will be for the Daubechies wavelet D_6 . Comparable results for the wavelets D_4 and D_8 are presented in appendix B.

Recall from the previous subsection that in order to calculate the coefficients $\{c_l\}_{l=0}^{M-1}$ the moments of the scaling function $\phi(x)$ must first be known. Let M_l be the l^{th} moment of the scaling function $\phi(x)$,

$$M_l = \int \phi(x)x^l dx, \quad (37)$$

and let μ_l be the l^{th} moment of the filter h_k ,

$$\mu_l = \sum_k k^l h_k. \quad (38)$$

The zero-th moment, M_0 , of $\phi(x)$ is 1 by the normalization of $\phi(x)$:

$$M_0 = \int \phi(x)dx = 1. \quad (39)$$

The zero-th moment of the coefficients h_k is found by integrating the dilation equation which defines $\phi(x)$:

$$\int \phi(x)dx = \sum_k h_k \int \phi(2x - k)dx, \quad (40)$$

and let $y = 2x - k$ to get,

$$1 = \frac{1}{2} \sum_k h_k \int \phi(y)dy, \quad (41)$$

which implies,

$$\mu_0 = \sum_k h_k = 2. \quad (42)$$

That is, the zero-th moments M_0 and μ_0 are the same for all Daubechies wavelets. Higher moments for μ_l can be found by straight-forward calculation using the coefficients provided by Daubechies [2]. The higher moments, M_l for $l > 0$, for the scaling function can be found from the following equation which is derived in appendix B:

$$M_m = \left(\frac{1}{2}\right)^{m+1} \sum_l \binom{m}{l} \mu_{m-l} M_l. \quad (43)$$

For the current example only the moments M_0 , M_1 , and M_2 are needed: $M_0 = 1$, $M_1 = \frac{1}{2}\mu_1$, and $M_2 = \frac{1}{6}((\mu_1)^2 + \mu_2)$. Given these three moments the coefficients $\{c_l\}_{l=0}^{M-1}$ can be found from

$$\sum_{l=0}^{M-1} l^m c_l = \int x^m \phi(x)dx \quad (44)$$

i	M_i	μ_i	c_i
0	1	2	.1080
1	.8174	1.6348	.9667
2	.6681	1.3363	-.0746

Table 1: Scaling function and filter moments for Daubechies 6 wavelets.

for $m = 0, 1, \dots, M - 1$. Specifically, for the D_6 wavelet the linear system in matrix form is,

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 1 & 4 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} M_0 \\ M_1 \\ M_2 \end{pmatrix}, \quad (45)$$

which has the solution $c_0 = .1080$, $c_1 = .9667$, and $c_2 = -.0746$. In tabular form, the complete results for D_6 are,

Recall that the quadrature formula used to approximate the scaling function has the form,

$$\sigma_k^0 = \sum_{l=0}^{M-1} c_l f(l+k). \quad (46)$$

If the function $f(x)$ is periodic then in matrix notation the above operation is $\vec{\sigma}_0 = C \vec{f}$ where C for D_6 and on a grid of 6 points is,

$$C = \begin{pmatrix} .108 & .967 & -.075 & 0 & 0 & 0 \\ 0 & .108 & .967 & -.075 & 0 & 0 \\ 0 & 0 & .108 & .967 & -.075 & 0 \\ 0 & 0 & 0 & .108 & .967 & -.075 \\ -.075 & 0 & 0 & 0 & .108 & .967 \\ .967 & -.075 & 0 & 0 & 0 & .108 \end{pmatrix}. \quad (47)$$

The important point here is that the above matrix is circulant. The ramifications of circularity are very important for the thesis of the paper. The definition of circulant matrices and the properties that they are imbued with is the subject of the next subsection.

3.3 Circulant Matrices

Strang [7] defines a circulant matrix as a constant-diagonal matrix which is ‘periodic, since the lower diagonals fold around to appear again as the upper diagonals.’ A thorough discussion of circulant matrices is given by Davis [8]. Circulant matrices

have the wonderful property that they can all be diagonalized by the same matrix, the Fourier matrix: An $N \times N$ Fourier matrix has as its ij -th element the entry $w^{(i-1)(j-1)}$ where $w^N = 1$. The most important ramification for this paper is that matrices which can be diagonalized by the same matrix commute. That is, the matrix C from the previous subsection will commute with the matrix R^0 which will be derived in section (1.4).

In general, circulant matrices arise whenever one is performing the matrix version of periodic discrete convolution. In numerical analysis periodic discrete convolution arises whenever one differentiates the evenly-spaced samples of a periodic equation which has constant coefficients. Let us be a bit more precise and illustrate how the operation of periodic discrete convolution yields a circulant matrix by stating the following theorem:

Theorem: A finite-length filter of length M applied to N evenly-spaced samples of periodic function, where $N > M$, will in matrix form yield a circulant matrix.

Proof: First of all, let the notation remain as above: c_0, c_1, \dots, c_{M-1} will represent the finite-length filter and f_0, f_1, \dots, f_{N-1} will represent the evenly-spaced samples of one period of the periodic function $f(x)$. Of course, the samples of $f(x)$ are periodic with period N . The application of the filter \vec{c} on the samples of $f(x)$ is the convolution:

$$\sigma_k = \sum_{l=0}^{M-1} c_l \hat{f}_{k-l},$$

where \hat{f}_i is the renaming of the elements of f_i so that the previous convolution is the same as the following expression. Furthermore, keep in mind that \hat{f}_i and f_i are periodic with period N .

$$\sigma_k = \sum_{l=0}^{M-1} c_l f_{l+k}.$$

Using the modulus function to keep the indices of f_i within one period, i.e., keep

$0 \leq i \leq N - 1$, the above equation can be written as,

$$\sigma_k = \sum_{l=0}^{M-1} c_l f_{\text{mod}(l+k, N)}.$$

Now, shift the indices by letting $j = l + k$ to get,

$$\sigma_k = \sum_{j=k}^{k+(M-1)} c_{j-k} f_{\text{mod}(j, N)}.$$

If the length- M filter \vec{c} is now 'padded' at the end with zeros so that it is now a length- N filter then the above equation can be rewritten as,

$$\sigma_k = \sum_{j=0}^{N-1} c_{\text{mod}(j-k, N)} f_j.$$

This is exactly a matrix multiply $\vec{\sigma} = C \vec{f}$ where the ij - *th* element of the matrix C is $c_{\text{mod}(j-i, N)}$, and this is the definition of a circulant matrix. This completes the proof.//

Note that the difference between a circulant matrix and a Toeplitz matrix is the wrapping around effect of the diagonals introduced by the use of the modulus function for the circulant matrix. That is, a circulant matrix is a special case of a Toeplitz matrix where the constant diagonals are periodic.

Before leaving the discussion on circulant matrices let one more interpretation be noted: to say that circulant matrices commute is to simply restate the important result from signal analysis that convolutions commute. That is, if one has two sequences c and r , which in the current scenario are periodic, then the order of convolution does not matter. This is easily proved with the Fourier transform:

$$\widehat{c * r} = \hat{c} \hat{r} = \hat{r} \hat{c} = \widehat{r * c}, \quad (48)$$

where \hat{r} denotes the Fourier transform of r and $*$ denotes periodic convolution. For this paper, c , of course, would be the quadrature operator and r would be the scaling function derivative operator which is the subject of section (4). In matrix notation

equation (48) is nothing more than,

$$C \cdot R^0 = R^0 \cdot C \tag{49}$$

In this section a quadrature formula has been found to approximate the scaling function coefficients of a given function, $f(x)$. In matrix form this quadrature formula leads to a circulant matrix assuming $f(x)$ is periodic. In the next section the wavelet derivative operator will be derived, and it will be shown that, once again, the assumption that $f(x)$ is periodic leads to an operator which in matrix form is circulant.

4 Derivative based on Wavelets

In the previous section the mapping from evenly-spaced samples of a periodic function, $f(x)$, to the approximate scaling function coefficients on the finest scale, σ_k^0 , was given. Recall that σ_k^j denotes the approximation to the exact scaling function coefficient s_k^j at scale j and position k . The mapping is nothing more than a quadrature formula which is exact when $f(x)$ is equal to a polynomial up to order $M - 1$, where M is the number of vanishing moments of the wavelet. The question now is how does one represent the derivative of $f(x)$ in the wavelet basis given that the wavelet expansion of $f(x)$ is already given.

The answer is given in the following subsections:

§4.1) A function $f(x)$ will be expanded in a wavelet basis and the expansion will be differentiated.

§4.2) The results from Beylkin [9] on derivative projections will be given.

§4.3) First, it will be noted that one can differentiate a wavelet expansion at any level of a wavelet decomposition and achieve the same derivative. Second, explicit wavelet decomposition will be performed accompanied by the appropriate differentiation matrix.

§4.4) The similarity between the wavelet-based derivative coefficients and finite difference derivative coefficients will be noted, and it will be shown that when one differentiates the wavelet expansion of a periodic function that the effect on the original function samples is equal to finite difference differentiation.

4.1 Expansion in a Wavelet Basis

The goal now is to find the wavelet and scaling function expansion of a periodic function $f(x)$. Given $f(x) \in L^2(R)$ one first projects onto the arbitrarily chosen finest scale $j = 0$ of the scaling function $\phi_k^0(x)$ which generates the space V_0 , i.e., let

P_{V_0} be the projection from the space $L^2(\mathbb{R})$ to the space V_0 , $P_{V_0} : L^2(\mathbb{R}) \rightarrow V_0$:

$$P_{V_0}f(x) = \sum_{k=0}^{N-1} s_k^0 \phi_k^0(x), \quad (50)$$

where due to the orthonormality of ϕ_k^0 over k in V_0 ,

$$s_k^0 = \int_{-\infty}^{\infty} f(x) \phi_k^0(x) dx. \quad (51)$$

Note that in the introduction the projection denoted by P_N would be the same as P_{V_0} using notation that is more amenable to wavelets. The derivative of $P_{V_0}f(x)$ is,

$$\frac{d}{dx} P_{V_0}f(x) = \sum_{k=0}^{N-1} s_k^0 \phi_k^{\prime 0}(x). \quad (52)$$

Of course, $\frac{d}{dx} P_{V_0}f(x)$ is not in V_0 and must be projected onto V_0 . First define the inner product $\langle f, g \rangle$ on $L^2(\mathbb{R})$ by

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g(x)dx. \quad (53)$$

Now the projection of $\frac{d}{dx} P_{V_0}f(x)$ onto V_0 is,

$$P_{V_0} \frac{d}{dx} P_{V_0}f(x) = \sum_{l=0}^{N-1} \langle \frac{d}{dx} P_{V_0}f, \phi_l^0 \rangle \phi_l^0(x), \quad (54)$$

or,

$$P_{V_0} \frac{d}{dx} P_{V_0}f(x) = \sum_{l=0}^{N-1} \sum_{k=0}^{N-1} s_k^0 \langle \phi_k^{\prime 0}, \phi_l^0 \rangle \phi_l^0(x). \quad (55)$$

The inner product $\langle \phi_k^{\prime 0}, \phi_l^0 \rangle$ is one of the results provided in [10].

In the previous paragraph $f(x)$ was expanded in a scaling-function expansion at the finest scale $j = 0$. Now $f(x)$ will be expanded in terms of scaling functions and wavelets at scale $j = 1$. Recall that $V_0 = V_1 \oplus W_1$. Now one must project from $L^2(\mathbb{R})$ onto V_1 and from $L^2(\mathbb{R})$ onto W_1 . Let both projections be denoted simultaneously by $P_{V_1 \oplus W_1}$. That is, $P_{V_1 \oplus W_1} : L^2(\mathbb{R}) \rightarrow V_1 \oplus W_1$. Let $P_{V_1 \oplus W_1}f(x)$ be the projection of $f(x)$ on $V_1 \oplus W_1$. Therefore, the expansion for $P_{V_1 \oplus W_1}f(x)$ is,

$$P_{V_1 \oplus W_1}f(x) = \sum_{k=0}^{N/2-1} s_k^1 \phi_k^1(x) + \sum_{k=0}^{N/2-1} d_k^1 \psi_k^1(x), \quad (56)$$

where due to the orthonormality of the basis functions $\phi_k^1(x)$ and $\psi_k^1(x)$ the coefficients s_k^1 and d_k^1 are given by

$$s_k^1 = \int_{-\infty}^{\infty} f(x)\phi_k^1(x)dx, \quad (57)$$

and

$$d_k^1 = \int_{-\infty}^{\infty} f(x)\psi_k^1(x)dx. \quad (58)$$

The derivative of $P_{V_1 \oplus W_1} f(x)$ is

$$\frac{d}{dx} P_{V_1 \oplus W_1} f(x) = \sum_{k=0}^{N/2-1} s_k^1 \phi_k^1(x) + \sum_{k=0}^{N/2-1} d_k^1 \psi_k^1(x). \quad (59)$$

Once again, the derivative of $P_{V_1 \oplus W_1} f(x)$ does not belong to $V_1 \oplus W_1$, and must, therefore, be projected back onto this space. The projection is,

$$P_{V_1 \oplus W_1} \frac{d}{dx} P_{V_1 \oplus W_1} f(x) = \quad (60)$$

$$\begin{aligned} & \sum_{l=0}^{N/2-1} \sum_{k=0}^{N/2-1} s_k^1 \langle \phi_k^1, \phi_l^1 \rangle \phi_l^1(x) \\ & + \sum_{l=0}^{N/2-1} \sum_{k=0}^{N/2-1} s_k^1 \langle \phi_k^1, \psi_l^1 \rangle \psi_l^1(x) \\ & + \sum_{l=0}^{N/2-1} \sum_{k=0}^{N/2-1} d_k^1 \langle \psi_k^1, \phi_l^1 \rangle \phi_l^1(x) \\ & + \sum_{l=0}^{N/2-1} \sum_{k=0}^{N/2-1} d_k^1 \langle \psi_k^1, \psi_l^1 \rangle \psi_l^1(x). \end{aligned}$$

The four inner products $\langle \phi_k^1, \phi_l^1 \rangle$, $\langle \phi_k^1, \psi_l^1 \rangle$, $\langle \psi_k^1, \phi_l^1 \rangle$, and $\langle \psi_k^1, \psi_l^1 \rangle$ are the key to finding the derivative of a wavelet expansion, and are provided in [9]. An outline of the derivation of these inner products is given in the next section.

4.2 Wavelet Coefficients of the Derivative

An arbitrary wavelet expansion of a function might contain wavelet coefficients and scaling coefficients at many scales. In [9] the projection coefficients that map from scaling function coefficients and wavelet function coefficients at a given scale to the

derivative scaling function coefficients and wavelet function coefficients at the same scale are derived. The matrix elements of these projections are computed from,

$$2^{-j}a_{i-l} = a_{il}^j = 2^{-2j} \int_{-\infty}^{\infty} \psi(2^{-j}x - i)\psi'(2^{-j}x - l)dx, \quad (61)$$

$$2^{-j}b_{i-l} = b_{il}^j = 2^{-2j} \int_{-\infty}^{\infty} \psi(2^{-j}x - i)\phi'(2^{-j}x - l)dx, \quad (62)$$

$$2^{-j}c_{i-l} = c_{il}^j = 2^{-2j} \int_{-\infty}^{\infty} \phi(2^{-j}x - i)\psi'(2^{-j}x - l)dx, \quad (63)$$

$$2^{-j}r_{i-l} = r_{il}^j = 2^{-2j} \int_{-\infty}^{\infty} \phi(2^{-j}x - i)\phi'(2^{-j}x - l)dx. \quad (64)$$

Since the above projections are always at a fixed scale, j , the projection coefficients are simply,

$$a_l = \int_{-\infty}^{\infty} \psi(x - l) \frac{d}{dx} \psi(x) dx, \quad (65)$$

$$b_l = \int_{-\infty}^{\infty} \psi(x - l) \frac{d}{dx} \phi(x) dx, \quad (66)$$

$$c_l = \int_{-\infty}^{\infty} \phi(x - l) \frac{d}{dx} \psi(x) dx, \quad (67)$$

$$r_l = \int_{-\infty}^{\infty} \phi(x - l) \frac{d}{dx} \phi(x) dx, \quad (68)$$

for $l \in \mathbb{Z}$. Furthermore, using the dilation equation which defines $\phi(x)$, $\phi(x) = \sum_k h_k \phi(2x - k)$, and the equation which defines $\psi(x)$, $\psi(x) = \sum_k g_k \phi(2x - k)$, the first three of the above four equations become,

$$a_i = \sum_{k=0}^{L-1} \sum_{l=0}^{L-1} g_k g_l r_{2i+k-l} \quad (69)$$

$$b_i = \sum_{k=0}^{L-1} \sum_{l=0}^{L-1} g_k h_l r_{2i+k-l} \quad (70)$$

$$c_i = \sum_{k=0}^{L-1} \sum_{l=0}^{L-1} h_k g_l r_{2i+k-l}. \quad (71)$$

It is apparent from the above equations that the coefficients r_l contain all the information concerning the derivative. The coefficients r_l can be found [9] from solving the following system of linear algebraic equations:

$$r_l = 2(r_{2l} + \frac{1}{2} \sum_{k=1}^{L/2} \alpha_{2k-1} (r_{2l-2k+1} + r_{2l+2k-1})), \quad (72)$$

and

$$\sum_l l r_l = -1, \quad (73)$$

where

$$\alpha_n = 2 \sum_{i=0}^{L-1-n} h_i h_{i+n}, \quad (74)$$

for $n = 1, \dots, L - 1$. The proof of the above proposition can be found in [9].

This section has given a brief outline of the derivation of the wavelet derivative projection coefficients. It is important to note that all the information for the wavelet derivative is contained in the coefficients $\{r_l\}$, and this point will be explored more in next section.

4.3 Derivative at Scale Zero of Scaling Function Only

Wavelet derivatives can be calculated at any level of a wavelet decomposition. The result will, of course, always be the same. That is, recall the relation from §(2), $V_j = V_{j+1} \oplus W_{j+1}$. As stated before, it is the convention of this paper to let V_0 represent the finest scale. Using the above relation, one could decompose V_0 any number of times. One decomposition yields $V_0 = W_1 \oplus V_1$, and a second decomposition yields $V_0 = W_1 \oplus W_2 \oplus V_2$. One could calculate the wavelet derivative in any one of these spaces. Once again, the goal of this paper is to understand the essence of a wavelet derivative, and since the derivative is the same regardless of the decomposition of the space, one should choose the simplest approach and calculate the derivative at scale $j = 0$ using only the scaling function coefficients.

This subsection contains four parts:

1. New notation will be introduced.
2. Wavelet decompositions and differentiation matrices will be given for the space V_0 as well as comments on data compression in this space.
3. Wavelet decompositions and differentiation matrices will be given for the space $W_1 \oplus V_1$ as well as comments on data compression in this space.

4. Wavelet decompositions and differentiation matrices will be given for the space $W_1 \oplus W_2 \oplus W_3 \oplus V_3$ as well as comments on data compression.

4.3.1 New Notation

To simplify the presentation, matrix notation will be used whenever possible. Begin by defining the matrix version of equations (29) and (30). Recall that these equations are

$$s_k^j = \sum_{n=1}^{n=2M} h_n s_{n+2k-2}^{j-1},$$

and

$$d_k^j = \sum_{n=1}^{n=2M} g_n s_{n+2k-2}^{j-1}.$$

Denote the decomposition matrix embodied by these two equations by $P_{N \times N}^{j,j+1}$ where the matrix subscripts denote the size of the matrix, and the superscripts indicate that P is decomposing from scaling function coefficients at scale j to scaling function and wavelet function coefficients at scale $j+1$. As before, let \vec{s}_j contain the scaling function coefficients at scale j . (Note: When vector notation is used the scale is given as a subscript, otherwise the location k is the subscript and the scale is the superscript.) P therefore maps \vec{s}_j onto \vec{s}_{j+1} and \vec{d}_{j+1} :

$$P_{N \times N}^{j,j+1} : \begin{bmatrix} \vec{s}_j \end{bmatrix} \rightarrow \begin{bmatrix} \vec{s}_{j+1} \\ \vec{d}_{j+1} \end{bmatrix}. \quad (75)$$

Note that the vectors at scale $j+1$ are half as long as the vectors at scale j . To illustrate further, suppose the wavelet being used is the four coefficient D_4 wavelet, and suppose one wants to project from 8 scaling function coefficients at scale j to 4 scaling function coefficients at scale $j+1$ and 4 wavelet coefficients at scale $j+1$.

The decomposition matrix in this case is,

$$P_{8 \times 8}^{j,j+1} \equiv \begin{bmatrix} h_1 & h_2 & h_3 & h_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_1 & h_2 & h_3 & h_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_1 & h_2 & h_3 & h_4 \\ h_3 & h_4 & 0 & 0 & 0 & 0 & h_1 & h_2 \\ g_1 & g_2 & g_3 & g_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_1 & g_2 & g_3 & g_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & g_1 & g_2 & g_3 & g_4 \\ g_3 & g_4 & 0 & 0 & 0 & 0 & g_1 & g_2 \end{bmatrix}. \quad (76)$$

Other decomposition matrices of different sizes will have the same structure as the above matrix.

For a bit more matrix notation, let the four matrices $A_{N \times N}^j$, $B_{N \times N}^j$, $C_{N \times N}^j$, and $R_{N \times N}^j$ contain the derivative projection coefficients defined in §(4.2) where, again, the subscripts denote the size of the matrix and the superscript denotes the scale on which the derivative projection is acting. The elements of the four matrices are,

$$A \leftrightarrow a_{ij} = a_{i-j},$$

$$B \leftrightarrow b_{ij} = b_{i-j},$$

$$C \leftrightarrow c_{ij} = c_{i-j},$$

and

$$R \leftrightarrow r_{ij} = r_{i-j},$$

and the mappings performed by the matrices are,

$$A^j : \vec{d}_j \rightarrow \vec{d}_j,$$

$$B^j : \vec{s}_j \rightarrow \vec{d}_j,$$

$$C^j : \vec{d}_j \rightarrow \vec{s}_j,$$

$$R^j : \vec{s}_j \rightarrow \vec{s}_j,$$

where \vec{s}_j and \vec{d}_j , as before, define the scaling and wavelet coefficients at scale j , and $\vec{\tilde{s}}_j$ and $\vec{\tilde{d}}_j$ denote the coefficients of the expansion of the derivative of a function which is initially defined by an expansion in \vec{s}_j and \vec{d}_j .

This concludes the new notation.

4.3.2 Wavelet Expansion and Derivative in V_0

As stated previously, one can calculate the derivative of a wavelet expansion at any level in the wavelet decomposition. This subsection will explore the first of three of the options. To be explicit, suppose that a periodic function $f(x)$ has been approximated on a grid with 16 scaling function coefficients to get \vec{s}_0 , and for the current argument assume that the coefficients have been calculated exactly, i.e., the notation \vec{s}_0 will be used instead of $\vec{\sigma}_0$. Furthermore due to the periodicity of $f(x)$ the coefficients \vec{s}_0 will also be periodic. The coefficients of the expansion of $\frac{d}{dx}f(x)$ in V_0 are found from \vec{s}_0 by an application of the previously defined matrix $R_{16 \times 16}^0$:

$$\begin{bmatrix} s_1^0 \\ s_2^0 \\ s_3^0 \\ s_4^0 \\ s_5^0 \\ s_6^0 \\ s_7^0 \\ s_8^0 \\ s_9^0 \\ s_{10}^0 \\ s_{11}^0 \\ s_{12}^0 \\ s_{13}^0 \\ s_{14}^0 \\ s_{15}^0 \\ s_{16}^0 \end{bmatrix} \xrightarrow{R_{16 \times 16}^0} \begin{bmatrix} s_1^{\prime 0} \\ s_2^{\prime 0} \\ s_3^{\prime 0} \\ s_4^{\prime 0} \\ s_5^{\prime 0} \\ s_6^{\prime 0} \\ s_7^{\prime 0} \\ s_8^{\prime 0} \\ s_9^{\prime 0} \\ s_{10}^{\prime 0} \\ s_{11}^{\prime 0} \\ s_{12}^{\prime 0} \\ s_{13}^{\prime 0} \\ s_{14}^{\prime 0} \\ s_{15}^{\prime 0} \\ s_{16}^{\prime 0} \end{bmatrix}, \quad (77)$$

which completely defines the derivative of $f(x)$ in the scaling function basis at scale $j = 0$.

For data compression purposes, the space V_0 is not a good space to work in. That is, the coefficients \vec{s}_0 represent the equivalent of a local averages. In a wavelet basis, it is often true that the coefficients of local high-frequency oscillations are small and can be set to zero without altering the character of the function being represented, but the coefficients of local averages usually represent essential information.

4.3.3 Wavelet Expansion and Derivative in $W_1 \oplus V_1$

Consider now a decomposition of the vector of scaling function coefficients \vec{s}_0 onto the scaling function and wavelet coefficients at scale $j = 1$ by an application of the matrix $P_{16 \times 16}^{0,1}$:

$$\begin{bmatrix} s_1^0 \\ s_1^0 \\ s_2^0 \\ s_2^0 \\ s_3^0 \\ s_3^0 \\ s_4^0 \\ s_4^0 \\ s_5^0 \\ s_5^0 \\ s_6^0 \\ s_6^0 \\ s_7^0 \\ s_7^0 \\ s_8^0 \\ s_8^0 \\ s_9^0 \\ s_9^0 \\ s_{10}^0 \\ s_{10}^0 \\ s_{11}^0 \\ s_{11}^0 \\ s_{12}^0 \\ s_{12}^0 \\ s_{13}^0 \\ s_{13}^0 \\ s_{14}^0 \\ s_{14}^0 \\ s_{15}^0 \\ s_{15}^0 \\ s_{16}^0 \end{bmatrix} \xrightarrow{P_{16 \times 16}^{0,1}} \begin{bmatrix} s_1^1 \\ s_1^1 \\ s_2^1 \\ s_2^1 \\ s_3^1 \\ s_3^1 \\ s_4^1 \\ s_4^1 \\ s_5^1 \\ s_5^1 \\ s_6^1 \\ s_6^1 \\ s_7^1 \\ s_7^1 \\ s_8^1 \\ d_1^1 \\ d_2^1 \\ d_3^1 \\ d_4^1 \\ d_5^1 \\ d_6^1 \\ d_7^1 \\ d_8^1 \end{bmatrix}, \quad (78)$$

As before, we have 16 basis functions in our space which is now $V_1 \oplus W_1$ rather than V_0 . In order to calculate the coefficients of the derivative expansion in $V_1 \oplus W_1$ the following projections are calculated:

$$\vec{s}_1 = R_{8 \times 8}^1 \cdot \vec{s}_1 + C_{8 \times 8}^1 \cdot \vec{d}_1, \quad (79)$$

and

$$\vec{d}_1 = A_{8 \times 8}^1 \cdot \vec{d}_1 + B_{8 \times 8}^1 \cdot \vec{s}_1, \quad (80)$$

where A , B , C , and R were all defined in the previous subsection. A more concise way to represent the derivative projections is in matrix notation:

$$\begin{bmatrix} \vec{s}_1 \\ \vec{d}_1 \end{bmatrix} = \begin{bmatrix} R_{8 \times 8}^1 & C_{8 \times 8}^1 \\ B_{8 \times 8}^1 & A_{8 \times 8}^1 \end{bmatrix} \cdot \begin{bmatrix} \vec{s}_1 \\ \vec{d}_1 \end{bmatrix}. \quad (81)$$

If one now applies the matrix $(P_{16 \times 16}^{0,1})^T$ (T denotes transpose and hence inverse for this unitary matrix) to the derivative coefficients at scale $j = 1$ then one gets the

derivative coefficients at scale $j = 0$:

$$\begin{bmatrix} \vec{s}_0 \end{bmatrix} = (P_{16 \times 16}^{0,1})^T \cdot \begin{bmatrix} \vec{s}_1 \\ \vec{d}_1 \end{bmatrix}, \quad (82)$$

and one gets exactly the same coefficients as before when the matrix $R_{16 \times 16}^0$ was applied to \vec{s}_0 . To emphasize, the derivative calculated at scale $j = 0$ and the derivative calculated at $j = 1$ yield exactly the same result. The importance of this observation is that in order to understand the essence of the wavelet derivative one need only be concerned with the action of the matrix $R_{N \times N}^0$ on the vector \vec{s}_0 .

For data compression the space $W_1 \oplus V_1$ is a fair space to work in. The coefficients \vec{s}_1 of the basis functions in V_1 represent local averages just as the coefficients of the basis functions in the space V_0 do. However, the basis functions in V_1 have broader support than the basis functions in V_0 and therefore represent averages over a larger amount of data (twice as much data to be exact). Therefore, once again the coefficients \vec{s}_1 usually carry essential information. The coefficients \vec{d}_1 of the basis functions in the space W_1 , on the other hand, carry information concerning local oscillations. That is, if the function being represented, $f(x)$, is globally smooth then the coefficients \vec{d}_1 will be near zero and can be set exactly to zero without altering the character of $f(x)$. In the solution of nonlinear partial differential equations where a sharp gradient, or shock, can develop, the coefficients \vec{d}_1 away from the shock would be close to zero whereas the coefficients near the shock would be large. Therefore, representing a function in $W_1 \oplus V_1$ is more versatile than simply staying in the space V_0 . Versatility continues to be enhanced as one decomposes into more and more wavelet subspaces as in the next and final scenario.

4.3.4 Wavelet Expansion and Derivative in $W_1 \oplus W_2 \oplus W_3 \oplus V_3$

Up to now our basis functions have all been at the same scale, i.e., initially our basis functions were contained in V_0 , and in the second scenario the basis functions were contained in V_1 and W_1 . In this subsection, however, the basis functions will

be contained in spaces at three different scales: W_1 , W_2 , W_3 , and V_3 . The full set of coefficients in this case and all the appropriate decompositions leading to these coefficients are,

$$\begin{bmatrix} s_1^0 \\ s_2^0 \\ s_3^0 \\ s_4^0 \\ s_5^0 \\ s_6^0 \\ s_7^0 \\ s_8^0 \\ s_9^0 \\ s_{10}^0 \\ s_{11}^0 \\ s_{12}^0 \\ s_{13}^0 \\ s_{14}^0 \\ s_{15}^0 \\ s_{16}^0 \end{bmatrix} \xrightarrow{P_{16 \times 16}^{0,1}} \begin{bmatrix} s_1^1 \\ s_2^1 \\ s_3^1 \\ s_4^1 \\ s_5^1 \\ s_6^1 \\ s_7^1 \\ s_8^1 \\ d_1^1 \\ d_2^1 \\ d_3^1 \\ d_4^1 \\ d_5^1 \\ d_6^1 \\ d_7^1 \\ d_8^1 \end{bmatrix} \xrightarrow{P_{8 \times 8}^{1,2}} \begin{bmatrix} s_1^2 \\ s_2^2 \\ s_3^2 \\ s_4^2 \\ d_1^2 \\ d_2^2 \\ d_3^2 \\ d_4^2 \\ d_1^1 \\ d_2^1 \\ d_3^1 \\ d_4^1 \\ d_5^1 \\ d_6^1 \\ d_7^1 \\ d_8^1 \end{bmatrix} \xrightarrow{P_{4 \times 4}^{2,3}} \begin{bmatrix} s_1^3 \\ s_2^3 \\ d_1^3 \\ d_2^3 \\ d_1^2 \\ d_2^2 \\ d_3^2 \\ d_4^2 \\ d_1^1 \\ d_2^1 \\ d_3^1 \\ d_4^1 \\ d_5^1 \\ d_6^1 \\ d_7^1 \\ d_8^1 \end{bmatrix}. \quad (83)$$

In matrix form the projection onto the coefficients of the derivative of the expansion is, where the matrix will be labeled M ,

$$M = \left[\begin{array}{c} \left[\begin{array}{cc} R_{2 \times 2}^3 & C_{2 \times 2}^3 \\ B_{2 \times 2}^3 & A_{2 \times 2}^3 \end{array} \right] P_{4 \times 4}^{2,3} C_{4 \times 4}^2 \\ B_{4 \times 4}^2 (P_{4 \times 4}^{2,3})^T A_{4 \times 4}^2 \\ \left[B_{8 \times 8}^1 (P_{8 \times 8}^{1,2})^T \left(\begin{array}{cc} P_{4 \times 4}^{2,3} & 0 \\ 0 & I \end{array} \right)^T \right] A_{8 \times 8}^1 \end{array} \right] \left[\begin{array}{c} \left(\begin{array}{cc} P_{4 \times 4}^{2,3} & 0 \\ 0 & I \end{array} \right) P_{8 \times 8}^{1,2} C_{8 \times 8}^1 \\ A_{8 \times 8}^1 \end{array} \right] \quad (84)$$

and \mathbf{M} performs the following mapping:

$$\mathbf{M} : \begin{bmatrix} s_1^3 \\ s_2^3 \\ d_1^3 \\ d_2^3 \\ d_1^2 \\ d_2^2 \\ d_3^2 \\ d_4^2 \\ d_1^1 \\ d_2^1 \\ d_3^1 \\ d_4^1 \\ d_5^1 \\ d_6^1 \\ d_7^1 \\ d_8^1 \end{bmatrix} \rightarrow \begin{bmatrix} s_1^3 \\ s_2^3 \\ d_1^3 \\ d_2^3 \\ d_1^2 \\ d_2^2 \\ d_3^2 \\ d_4^2 \\ d_1^1 \\ d_2^1 \\ d_3^1 \\ d_4^1 \\ d_5^1 \\ d_6^1 \\ d_7^1 \\ d_8^1 \end{bmatrix}. \quad (85)$$

For data compression, this is the most useful set of subspaces. The space now is represented as $W_1 \oplus W_2 \oplus W_3 \oplus V_3$. For the same reasons as before the coefficients of basis functions in the subspace V_3 cannot be ignored. It is likely, however, that the function $f(x)$ being represented is smooth in most of the domain allowing one to disregard the majority of the coefficients of the basis functions in the subspace $W_1 \oplus W_2 \oplus W_3$. In fact, it is more likely that the coefficients for the basis functions in W_1 will be negligible than for the coefficients for the basis functions in W_3 . This is because the basis functions in W_3 have larger support than the basis functions in W_2 and W_1 .

In summary, an attempt has been made to illustrate that the derivative coefficients of a scaling and wavelet expansion can be calculated at any scale. The proper set of wavelet subspaces depends on the problem at hand. The goal for this author is to understand exactly what wavelets are and what they are doing, therefore, scale $j = 0$, i.e., the space V_0 , provides the clearest scenario in which to work without sacrificing essential properties of wavelets.

Given, now, that it is sufficient to work on scale $j = 0$ to understand exactly

what the wavelet derivative does, one must understand the ramifications of applying the matrix R^0 to the vector \vec{s}_0 . In the next subsection the similarity between the above defined matrix R^0 and finite difference formulas for taking the derivative will be explored.

4.4 Wavelet Derivatives and Finite Difference

As the previous subsection illustrated, the essential properties of the wavelet derivative are contained in the elements of the matrix R . Recall that R is the matrix form of the mapping from \vec{s}_0 to \vec{s}'_0 . The surprising property that the matrix R exhibits is, however, that it can also differentiate evenly-spaced samples of a function. That is, R acts as a finite-difference operator when applied to the samples of a function.

This subsection is in three parts:

1. The similarity between wavelet derivative coefficients and finite difference coefficients is noted.
2. The finite difference accuracy of the coefficients $\{r_l\}$ derived in [9] will be illustrated, and it will be proved in general that the coefficients $\{r_l\}$ can differentiate polynomials exactly up through order $2M$ for coefficients $\{r_l\}$ that were derived for Daubechies wavelets D_{2M} .
3. In the finite element method under certain conditions one achieves a very high order of accuracy called 'superconvergence.' In wavelet differentiation a similar phenomenon is encountered. This phenomenon is defined and a short explanation is offered.

4.4.1 Finite Difference Coefficients

First of all, it is useful to simply note the similarity between the coefficients of centered finite difference formulas and the coefficients used to construct the matrix R . The

Order of Accuracy	Coefficients								
2			$-\frac{1}{2}$	0	$\frac{1}{2}$				
4		$\frac{1}{12}$	$-\frac{2}{3}$	0	$\frac{2}{3}$	$-\frac{1}{12}$			
6		$-\frac{1}{60}$	$\frac{3}{20}$	$-\frac{3}{4}$	0	$\frac{3}{4}$	$-\frac{3}{20}$	$\frac{1}{60}$	
8	$\frac{1}{280}$	$-\frac{4}{105}$	$\frac{1}{5}$	$-\frac{4}{5}$	0	$\frac{4}{5}$	$-\frac{1}{5}$	$\frac{4}{105}$	$-\frac{1}{280}$

Table 2: Optimal centered finite difference coefficients with order of accuracy.

Wavelet	Convolution Coefficients						
D_2	0	$\frac{1}{2}$					
D_4	0	$\frac{2}{3}$	$-\frac{1}{12}$				
D_6	0	$\frac{272}{365}$	$-\frac{53}{365}$	$\frac{16}{1095}$	$\frac{1}{2920}$		
D_8	0	$\frac{39296}{49553}$	$-\frac{76113}{396424}$	$\frac{1664}{49553}$	$-\frac{2645}{1189272}$	$-\frac{128}{743295}$	$\frac{1}{1189272}$

Table 3: Scaling function derivative convolution coefficients for Daubechies wavelets.

following is a table of centered finite difference coefficients and the order of accuracy of the approximation to the derivative:

Recall that the elements of the matrix R derived in [9] provide the transformation from scaling function coefficients of a function to the scaling function coefficients of the derivative of the same function. The elements of R for the D_2 and D_4 wavelet derivatives are, as is shown in the following table, exactly the same as the coefficients for the 2-nd and 4-th order centered finite difference formulas. Note that the wavelet filters become quite long with increasing order. Therefore, only the right side of the filter will be shown keeping in mind that these filters are antisymmetric:

The fractions for the D_6 and D_8 wavelets are exact but complicated and provide little insight. Compare the following decimal representations of the 6-th and 8-th order finite difference operators to the decimal representations of D_6 and D_8 . Once again, only the right-hand side of these antisymmetric filters is shown:

The coefficients for the D_6 and D_8 derivatives are not the same as the coefficients for the 6-th order and 8-th order centered finite difference derivatives, but the differences are not large. Surprisingly, however, D_6 has the same accuracy as the 6-th order finite difference operator, and D_8 has the same accuracy as the 8-th order finite

FD-6 and D_6 Coefficients			
FD-6	0 .750	-.150	.017
D_6	0 .745	-.145	.015 .0003

Table 4: Comparison between numerical values of optimal 6th order centered finite difference coefficients and Daubechies 6 scaling function derivative convolution coefficients.

FD-8 and D_8 Coefficients				
FD-8	0 .80	-.20	.038	-.0036
D_8	0 .79	-.19	.034	-.0022 -.0002 .0000008

Table 5: Comparison between numerical values of optimal 8th order centered finite difference coefficients and Daubechies 8 scaling function derivative convolution coefficients.

difference operator. §4.4.2 will establish this accuracy.

4.4.2 Finite Difference Accuracy

To establish the finite-difference accuracy of the wavelet-based differentiation coefficients note that a centered-finite-difference derivative approximation with $2K$ anti-symmetric, $r_k = -r_{-k}$ implying $r_0 = 0$, coefficients, $(r_k)_{k=-K}^K$, can be written

$$\dot{f}(x_j) \sim \sum_{k=1}^K r_k (f_{j+k} - f_{j-k}). \quad (86)$$

If the above equation is exact for $f(x) = x^q$ for $q = 0, \dots, N$ but not for $q = N + 1$ then the equation is said to be N -th order accurate. Therefore, one must check to see if

$$qx_j^{q-1} = \sum_{k=1}^K r_k (x_{j+k}^q - x_{j-k}^q), \quad (87)$$

when $f(x) = x^q$. To simplify, one can let $x_j = j$ and check the following:

$$qj^{q-1} = \sum_{k=1}^K r_k ((j+k)^q - (j-k)^q). \quad (88)$$

Now, treating the coefficients derived in [9] as nothing more than finite-difference coefficients one can check the accuracy. The following table contains the results of applying the coefficients from [9] to various polynomials:

Wavelet Derivative	Exact up to	But not for
D_2	x^2	x^3
D_4	x^4	x^5
D_6	x^6	x^7
D_8	x^8	x^9
D_{10}	x^{10}	x^{11}
D_{12}	x^{12}	x^{13}
D_{14}	x^{14}	x^{15}
D_{16}	x^{16}	x^{17}
D_{18}	x^{18}	x^{19}

Table 6: The degree of polynomials differentiated exactly by various Daubechies scaling function derivative coefficients.

The pattern in this table is obvious and leads to the following theorem:

Theorem: If $\phi(x)$ is the scaling function for the Daubechies wavelet denoted by D_{2M} , where M is the number of vanishing moments of the wavelet, then the coefficients $\{r_l\}$ derived from

$$r_l = \int_{-\infty}^{\infty} \phi(x-l) \frac{d}{dx} \phi(x) dx$$

and applied the to evenly-spaced samples of a function act as a finite difference derivative operator of order $2M$.

The proof of this theorem requires two results, as well as the Fourier Transform of $\phi(x)$. First $\hat{\phi}(\xi)$ will be found followed by the two results which are needed and stated in theorem form. Perhaps the first proof would suffice assuming the second result is well-known. However, to be complete the second result is also proved.

Fourier Transform of $\phi(x)$: Recall,

$$\phi(x) = \sqrt{2} \sum_{k=0}^{L-1} h_k \phi(2x-k).$$

Define the Fourier Transform of $\phi(x)$ as,

$$\hat{\phi}(\xi) = \int_{-\infty}^{\infty} \phi(x) e^{i\xi x} dx.$$

Therefore,

$$\hat{\phi}(\xi) = \sqrt{2} \sum_{k=0}^{L-1} h_k \int_{-\infty}^{\infty} \phi(2x - k) e^{i\xi k} dx.$$

Let $y = 2x - k$ which implies $dx = dy/2$ to get

$$\begin{aligned} &= \frac{1}{\sqrt{2}} \sum_{k=0}^{L-1} h_k \int_{-\infty}^{\infty} \phi(y) e^{i(\xi/2)(y+k)} dy \\ &= \frac{1}{\sqrt{2}} \sum_{k=0}^{L-1} h_k e^{i k \frac{\xi}{2}} \int_{-\infty}^{\infty} \phi(y) e^{i y \frac{\xi}{2}} dy, \end{aligned}$$

or simply

$$\hat{\phi}(\xi) = \hat{H}(\xi/2) \hat{\phi}(\xi/2),$$

recalling that $\hat{H}(\xi) = \frac{1}{\sqrt{2}} \sum_{k=0}^{L-1} h_k e^{i k \xi}$. Furthermore, we get $\hat{\phi}(\xi) = \hat{H}(\xi/2) \hat{H}(\xi/4) \hat{\phi}(\xi/4)$.

This implies,

$$\hat{\phi}(\xi) = \hat{\phi}(0) \prod_{j=1}^{\infty} \hat{H}\left(\frac{\xi}{2^j}\right),$$

but $\phi(x)$ is normalized, $\hat{\phi}(0) = \int \phi(x) dx = 1$. Therefore,

$$\hat{\phi}(\xi) = \prod_{j=1}^{\infty} \hat{H}\left(\frac{\xi}{2^j}\right).$$

Theorem: The Fourier Transform of $\{r_l\}$ is of the form,

$$\hat{r}(\xi) = i\xi + c\xi^{2M+1} + h.o.t.,$$

where $c \in \mathbb{C}$ is some constant, and 'h.o.t.' denotes higher-order terms and will be used again in the proof.

Proof: Begin with the expression for $\{r_l\}$:

$$\rho(y) = \int_{-\infty}^{\infty} \phi(x - y) \frac{d}{dx} \phi(x) dx.$$

If we define,

$$f(x) = \phi(-x),$$

and

$$g(x) = \frac{d}{dx}\phi(x),$$

then

$$\rho(y) = f * g(y)$$

where '*' is the convolution operator. The convolution theorem states that the Fourier Transform (continuous or discrete) of a convolution is the product of the Fourier Transforms:

$$\hat{\rho}(\xi) = \hat{f}(\xi)\hat{g}(\xi).$$

If we define r_l as,

$$r_l = \rho(l)$$

then the semi-discrete Fourier transform of r_l is

$$\hat{r}(\xi) = \sum_{k=-\infty}^{\infty} \rho(\xi + 2\pi k).$$

where,

$$\hat{\rho}(\xi) = \int_{-\infty}^{\infty} \rho(x)e^{ix\xi} dx,$$

and

$$\hat{r}(\xi) = \sum_{k=-\infty}^{\infty} r_k e^{ik\xi}.$$

Calculate the needed Fourier Transforms to get,

$$\hat{f}(\xi) = \overline{\hat{\phi}(\xi)},$$

where $\bar{\cdot}$ denotes conjugation, and

$$\hat{g}(\xi) = i\xi\hat{\phi}(\xi).$$

Combine these results to get the Fourier Transform of $\{\rho_l\}$:

$$\hat{\rho}(\xi) = |\hat{\phi}(\xi)|^2 i\xi.$$

Now, we need to know the behaviour of $|\hat{\phi}(\xi)|^2$. Recall from the definitions that,

$$\hat{H}(\xi) = (1 + e^{i\xi})^M \left(\frac{1}{2}\right)^M Q(e^{i\xi}),$$

where $Q(e^{i\xi})$ does not have poles or zeros at $\xi = \pi$, see [2]. That is, $\hat{H}(\xi)$ has a zero of order M at $\xi = \pi$. Therefore, $\hat{H}(\xi + \pi)$ has a zero of order M at $\xi = 0$, i.e.,

$$\hat{H}(\xi + \pi) = c\xi^M + h.o.t.,$$

then

$$|\hat{H}(\xi + \pi)|^2 = a\xi^{2M} + h.o.t.,$$

where $a = |c|^2$. Recall from the definitions that,

$$|\hat{H}(\xi)|^2 + |\hat{H}(\xi + \pi)|^2 = 1.$$

Combine the two previous relations to get,

$$|\hat{H}(\xi)|^2 = 1 - a\xi^{2M} + h.o.t.$$

That is,

$$\frac{d^n}{d\xi^n} |\hat{H}(\xi)|^2 |_{\xi=0} = 0,$$

for $n = 1, \dots, 2M - 1$. The Fourier Transform of $\phi(x)$ was found above:

$$\hat{\phi}(\xi) = \prod_{j=1}^{\infty} \hat{H}\left(\frac{\xi}{2^j}\right).$$

We get an expression for $|\hat{\phi}(\xi)|^2$ from,

$$\hat{\phi}(\xi) \overline{\hat{\phi}(\xi)} = \prod_{j=1}^{\infty} \hat{H}\left(\frac{\xi}{2^j}\right) \prod_{j=1}^{\infty} \overline{\hat{H}\left(\frac{\xi}{2^j}\right)},$$

or

$$|\hat{\phi}(\xi)|^2 = \prod_{j=1}^{\infty} |\hat{H}\left(\frac{\xi}{2^j}\right)|^2.$$

Now, derivatives of this expression have the form,

$$\frac{d^n}{d\xi^n} |\hat{\phi}(\xi)|^2 = \prod_{j=1}^{\infty} \frac{d^n}{d\xi^n} |\hat{H}\left(\frac{\xi}{2^j}\right)|^2,$$

and one can see that if, $\frac{d^n}{d\xi^n} |\hat{H}(\xi)|^2|_{\xi=0} = 0$, for $n = 1, \dots, 2M - 1$, then

$$\frac{d^n}{d\xi^n} |\hat{\phi}(\xi)|^2|_{\xi=0} = 0,$$

for $n = 1, \dots, 2M - 1$. From this information we can see that a series expansion of $|\hat{\phi}(\xi)|^2$ about $\xi = 0$ would be of the form,

$$|\hat{\phi}(\xi)|^2 = a + b\xi^{2M} + h.o.t.$$

But, $\phi(x)$ is normalized implying that $\hat{\phi}(0) = 1$ and therefore $|\hat{\phi}(0)|^2 = 1$. The expansion becomes,

$$|\hat{\phi}(\xi)|^2 = 1 + b\xi^{2M} + h.o.t.,$$

where $b \in C$. Recall that we are looking for the semi-discrete Fourier transform of $\hat{r}(\xi)$, which we see from above is,

$$\hat{r}(\xi) = \sum_{k=-\infty}^{\infty} \rho(\xi + 2\pi k) = \sum_{k=-\infty}^{\infty} i(\xi + 2\pi k) |\hat{\phi}(\xi + 2\pi k)|^2.$$

We now need to find the behaviour of $|\hat{\phi}(\xi + 2\pi k)|^2$ when $k \neq 0$. Note that in the expression,

$$|\hat{\phi}(\xi + 2\pi k)|^2 = \prod_{j=1}^{\infty} |\hat{H}(\frac{\xi + 2\pi k}{2^j})|^2,$$

the arguement,

$$\frac{\xi + 2\pi k}{2^j} = \frac{\xi}{2^j} + \frac{k\pi}{2^{j-1}},$$

will for some j be equal to

$$\frac{\xi + 2\pi k}{2^j} = \frac{\xi}{2^j} + \pi,$$

modulus 2π . That is, if k is odd in the expression $(k\pi)/(2^{j-1})$ then stop when $j = 1$ since we can subtract multiples of 2π from $(k\pi)$ without changing \hat{H} since \hat{H} is 2π periodic. If k is even, then for some j , the number $(k)/(2^{j-1})$ will be odd at which point we again subtract some multiple of 2π . Consequently, in the infinite product,

$$|\hat{\phi}(\xi + 2\pi k)|^2 = \prod_{j=1}^{\infty} |\hat{H}(\frac{\xi + 2\pi k}{2^j})|^2,$$

there will always be a term, when $k \neq 0$, on the right hand side with the form, $|\hat{H}(\frac{\xi}{2^j} + \pi)|^2$. But, from above, we found that,

$$|\hat{H}(\frac{\xi}{2^j} + \pi)|^2 = O(\xi^{2M}) + h.o.t.$$

But this implies that for $k \neq 0$ that the contributions to the infinite sum,

$$\sum_{k=-\infty}^{\infty} |\hat{\phi}(\xi + 2\pi k)|^2$$

are of $O(\xi^{2M})$. That is,

$$\sum_{k=-\infty}^{\infty} |\hat{\phi}(\xi + 2\pi k)|^2 = 1 + b\xi^{2M} + h.o.t.$$

Ultimately, we need the semi-discrete Fourier transform of $\{r\}$:

$$\hat{r}(\xi) = i \sum_{k=-\infty}^{\infty} (\xi + 2\pi k) |\hat{\phi}(\xi + 2\pi k)|^2,$$

or

$$\hat{r}(\xi) = i\xi \sum_{k=-\infty}^{\infty} |\hat{\phi}(\xi + 2\pi k)|^2 + 2\pi i \sum_{k=-\infty}^{\infty} k |\hat{\phi}(\xi + 2\pi k)|^2.$$

We already know the behaviour of the first term on the right-hand side. The second term on the right-hand side cannot contribute powers of ξ lower than $2M$ since it differs from the summation in the first term only by a multiple of k which does not allow the low power contribution when $k = 0$. The final step is to illustrate the second term on the right-hand side is an odd function implying that the lowest power of ξ it can contribute is $2M + 1$, the first odd number past $2M$. That is, define

$$f(\xi, k) = k |\hat{\phi}(\xi + 2\pi k)|^2,$$

and note that in the infinite summation that there is always a term with positive k and a term with negative k . The summation of all such $+k$ terms and $-k$ terms yields odd functions.

$$f(\xi, k) + f(\xi, -k) = f(-\xi, k) + f(-\xi, -k),$$

and this implies the desired result leaving us with the conclusion of the proof that,

$$\hat{r}(\xi) = i\xi + ib\xi^{2M+1} + h.o.t.$$

This completes the proof.//

Lemma: Let $\{r_l\}$ be a finite set of coefficients. These coefficients can be called the coefficients of a finite difference approximation to a first derivative, or these coefficients can be called a finite impulse response filter, or FIR filter. Furthermore, let the coefficients be antisymmetric: $r_l = -r_{-l}$ which implies $r_0 = 0$. If the Discrete Fourier Transform, or DFT, of $\{r_l\}$ is of the form

$$\hat{r}(\xi) = i\xi + c\xi^{m+1} + h.o.t., \quad (89)$$

for some constant $c \in C$, then the filter $\{r_l\}$ when applied to evenly-spaced samples of a function can differentiate in a finite difference sense with accuracy of order m . That is, $\{r_l\}$ can differentiate polynomials exactly up to x^m .

Before the proof, note that the DFT of a filter which is designed to approximate differentiation in the space domain should approximate $i\xi$ in the frequency domain:

$$\frac{d}{dx}e^{i\xi x} = i\xi e^{i\xi x}. \quad (90)$$

That is, differentiation filters are attempting to approximate the frequency of a pure sinusoidal mode.

Proof:

Let the DFT for $\{r_l\}$ be defined as,

$$\hat{r}(\xi) = \sum_l r_l e^{i\xi l}. \quad (91)$$

Break up the summation to write as,

$$\hat{r}(\xi) = r_0 + \sum_{l \geq 1} r_l e^{i\xi l} + \sum_{l \leq -1} r_l e^{i\xi l}. \quad (92)$$

Now, impose the antisymmetry to get,

$$\hat{r}(\xi) = \sum_{l \geq 1} r_l (e^{i\xi l} - e^{-i\xi l}). \quad (93)$$

Using the series expansion about zero for the complex exponential one gets,

$$\hat{r}(\xi) = \sum_{l \geq 1} r_l \left(\sum_{k=0}^{\infty} \frac{(i\xi l)^k}{k!} - \frac{(-i\xi l)^k}{k!} \right), \quad (94)$$

or factor to get,

$$\hat{r}(\xi) = \sum_{l \geq 1} r_l \sum_{k=0}^{\infty} \frac{(i\xi)^k}{k!} (l^k - (-l)^k). \quad (95)$$

Interchange the summations to get,

$$\hat{r}(\xi) = \sum_{k=0}^{\infty} \frac{(i\xi)^k}{k!} \sum_{l \geq 1} r_l (l^k - (-l)^k). \quad (96)$$

Recall that the hypothesis was that $\hat{r}(\xi) = i\xi + c\xi^{m+1} + h.o.t.$ This implies that,

$$0 = \sum_{l \geq 1} r_l (l^k - (-l)^k) \quad (97)$$

for $k = 0$ and for $2 \leq k \leq m$. Furthermore, $1 = \sum_{l \geq 1} r_l (l^k - (-l)^k)$ must hold when $k = 1$. But these are exactly the conditions that must hold for a filter to be able to differentiate exactly polynomials up through order m which are centered at zero. The proof for polynomials centered at some arbitrary position requires a shift in the index but the results are the same. This completes the proof.//

This completes the theorems which are at the heart of the paper. The next section discusses the high order of accuracy of the coefficients $\{r_l\}$.

4.4.3 'Superconvergence'

Note that the matrix R^0 can differentiate exactly polynomials up to degree $2M$ for the Daubechies wavelet D_{2M} when thought of as a finite-difference operator, even though the scaling function subspace V_0 can only represent exactly polynomials only up to degree $M - 1$. A Similar phenomenon is encountered in the finite element method

under particular choices of the approximation grid and is known as superconvergence [11].

To understand the source of superconvergence in the wavelet derivative it is sufficient to have a good understanding of the proofs in the previous subsection. Let us note the sources of the powers of ξ in the expression for the DFT of the coefficients $\{r_l\}$:

$$\hat{r}(\xi) = i\xi + c\xi^{2M+1} + h.o.t.$$

Recall the definition of $\{r_l\}$,

$$r_l = \int_{-\infty}^{\infty} \phi(x-l) \frac{d}{dx} \phi(x) dx,$$

as well as the definitions of $\phi(x)$ and $\frac{d}{dx} \phi(x)$:

$$\phi(x) = \sum_{l=0}^{L-1} \phi(2x-l),$$

$$\dot{\phi}(x) = 2 \sum_{l=0}^{L-1} \dot{\phi}(2x-l).$$

The sources of the powers of ξ are now apparent: M powers come from $\phi(x)$ and $M+1$ powers come from $\frac{d}{dx} \phi(x)$. The 'superconvergence' for the wavelet derivative can be explained by the similarity between the equations which define $\phi(x)$ and $\frac{d}{dx} \phi(x)$. That is, they are defined by dilation equations which differ only by a multiple of 2.

The next section will summarize and conclude this paper.

5 Conclusion

A restatement of the thesis of this paper is given first followed by a brief outline of the argument.

Given the evenly-spaced samples of a periodic function, \vec{f} , then the matrix R^0 derived for the Daubechies wavelet D_{2M} has the effect, when applied to \vec{f} , of a finite-difference derivative operator of degree $2M$.

The heart of the argument of this paper is contained in §(3) and §(4). In §(3) it was established that if given the evenly-spaced samples of a periodic function $f(x)$ then the scaling function coefficients \vec{s}_0 of the function at the finest scale can be approximated by a quadrature formula which in matrix form,

$$\vec{\sigma}_0 = C\vec{f},$$

yields a circulant matrix C , where $\vec{\sigma}_0$ approximates \vec{s}_0 . Furthermore, in §(3) it was noted that *all* circulant matrices with the same dimensions commute. In §(4) it was noted that the coefficients which map the scaling function coefficients at the finest scale of a periodic function to the scaling function coefficients at the finest scale of the derivative of the function is also circulant in form when written in matrix notation,

$$\vec{s}'_0 = R^0\vec{s}_0.$$

Furthermore, it was observed that the matrix R^0 can differentiate evenly-spaced samples of a polynomial in a finite-difference sense exactly up to the order of the wavelet. Also, when R^0 is applied to the evenly-spaced samples of a *periodic* function then R^0 is circulant. Now, combine the results of §(3) and §(4) to get the following relation:

$$\vec{f}' = C^{-1}R^0C\vec{f}.$$

Throughout the paper it has been noted that C and R^0 are circulant in form when $f(x)$ is periodic and that circulant matrices commute. Therefore, the previous relation simply becomes,

$$\vec{f}' = R^0\vec{f},$$

where R^0 is acting as a finite difference operator.

A note concerning notation is in order. In the introduction the matrices C , D , and the *differentiation matrix* \mathcal{D} were defined. Under the scenario developed in this paper, the wavelet matrix C is the same as the matrix C from the introduction. The matrix D from the introduction becomes the matrix R^0 . Likewise, the matrix \mathcal{D} is also R^0 since for wavelets C and R^0 commute. That is, for evenly-spaced samples and periodicity R^0 is the wavelet *differentiation matrix* which has the effect of a finite difference operator.

The importance of the thesis of this paper is that under periodicity and an evenly-spaced grid one can understand the wavelet differentiation matrix in terms of a finite difference operator with the accuracy given by the superconvergence theorem.

Acknowledgements

I would like to acknowledge the guidance of my thesis advisor Professor David Gottlieb at Brown University who asked the questions that led to this work.

References

- [1] E. Tadmor, "Stability of Finite-Difference, Pseudospectral and Fourier-Galerkin Approximations for Time-Dependent Problems," *SIAM Review*, Vol. 29, No. 4, December 1987.
- [2] I. Daubechies, "Orthonormal Basis of Compactly Supported Wavelets", *Comm. Pure Appl. Math.*, **41** (1988), pp. 909-996.
- [3] Esteban, D. and Garland, C., "Applications of Quadrature Mirror Filters to Split Band Voice Coding Schemes," *Proc. Int. Conf. Acoustic Speech and Signal Proc.*, May 1977.
- [4] S. Mallat, "Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(\mathbb{R})$ ", *Trans. Amer. Math. Soc.*, Vol. 315, No. 1, pp. 69-87, Sept. 1989.
- [5] S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", *IEEE Trans. Pattern Anal. and Machine Intel.*, 11 (1989), pp. 674-693.
- [6] G. Strang, "Wavelets and Dilation Equations: A Brief Introduction", *SIAM Review*, Vol. 31, No.4, pp. 614-627, Dec. 1989.
- [7] G. Strang, "Introduction to Applied Mathematics", *Wellesley-Cambridge Press*, 1986, pp. 297-298.
- [8] P. Davis, "Circulant Matrices", *Wiley-Interscience*, 1979.
- [9] G. Beylkin, "On the Representation of Operators in Bases of Compactly Supported Wavelets", *SIAM J. Num. Anal.*, 29(6): 1716-1740, 1992.
- [10] G. Beylkin, R. Coifman and V. Rokhin, "Fast Wavelet Transforms and Numerical Algorithms I", *Comm. Pure Appl. Math.*, **64** (1991), pp. 141-184.

- [11] G. Fairweather, "Finite Element Galerkin Methods for Differential Equations", Marcel Dekker, 1978, pp. 43-49.

Appendix A: Wavelets Supported on $(0, 3M)$

In this appendix our wavelets are supported on $[0, 3M]$ where M is the number of vanishing moments of the wavelet. These are not the usual Daubechies wavelets, but for these wavelets the scaling function coefficients of a periodic function $f(x)$ can be approximated with error of order M simply by sampling $f(x)$ at the correct location.

To begin, assume that there exist a unique τ_M , fixed for a fixed number of vanishing moments, M , of the wavelet, such that

$$\int \phi(x + \tau_M)x^m dx = 0$$

for $m = 1, 2, \dots, M - 1$. Furthermore, recall the definition of the scaling function coefficient and expand the integrand $f(x)$ in a Taylor series about x_0 ($f'_0 = f'(x_0)$):

$$s_k^0 = \int f(x)\phi(x - k)dx =$$

$$f_0 \int \phi(x - k)dx + f'_0 \int (x - x_0)\phi(x - k)dx + f''_0 \int (x - x_0)^2\phi(x - k)dx + \dots$$

Now, shift the variable of integration by $y = x - \tau - k$, and choose the point of expansion, x_0 , to be $\tau + k$ to get,

$$s_k^0 =$$

$$f(\tau + k) \int \phi(y + \tau)dy + f'(y + k) \int y\phi(y + \tau)dy + f''(y + k) \int y^2\phi(y + \tau)dy + \dots$$

Now, rename τ as τ_M and the above integrals are of the form,

$$\int \phi(x + \tau_M)x^m dx = 0,$$

and therefore vanish for $m = 1, \dots, M - 1$ leading to,

$$s_k^0 = f(\tau_M + k) + f^{(M)}(\tau_M + k) \int y^M \phi(y + \tau_M)dy + \dots,$$

i.e., the approximation of the scaling function coefficient s_k^0 up to order M is made by sampling $f(x)$ at the position $\tau_M + k$.

Note that all of the above calculations could have been carried out for the first derivative of $f(x)$ giving an approximation to the scaling function coefficients, s_k^0 , of $f'(x)$:

$$s_k^0 = f'(\tau + k) + f^{(M+1)}(\tau + k) \int y^M \phi(y + \tau) dy + \dots$$

It was assumed above that there exist one τ_M such that

$$\int \phi(x + \tau_M) x^m dx = 0,$$

for $m = 1, \dots, M - 1$. For $m = 1$ this τ_M is easy to find:

$$\begin{aligned} \int \phi(x + \tau_M) x dx &= \int \phi(x) (x - \tau_M) dx \\ &= \int x \phi(x) dx - \tau_M \int \phi(x) dx. \end{aligned}$$

But $\int \phi(x) dx = 1$, therefore,

$$\tau_M = \int x \phi(x) dx.$$

That is, τ_M is simply the first moment of $\phi(x)$. To find τ_M for $m > 1$ the calculations are simple but a bit longer and require the result from the following theorem to show that there is one τ_M which is the same for all $m = 1, \dots, M - 1$.

If $\int \phi(x) dx = 1$ and there exists τ such that $\int \phi(x + \tau) x^m dx = 0$ for $m = 1, \dots, M - 1$ then $\int \phi(x) x^m dx = (\int \phi(x) x dx)^m$ for $m = 1, \dots, M - 1$.

Proof: Start with

$$\int \phi(x + \tau) x^m dx = 0,$$

and let $y = x + \tau$ to get,

$$\int \phi(y) (y - \tau)^m dy = 0.$$

Using the binomial theorem this becomes,

$$\int \phi(y) \sum_{r=0}^m \binom{m}{r} y^r (-\tau)^{m-r} dy = 0.$$

Let the moments of $\phi(x)$ be denoted by $M_l = \int \phi(x)x^l dx$ to get

$$\sum_{r=0}^m \binom{m}{r} (-\tau)^{m-r} M_r = 0.$$

A simple calculation yields $\tau = M_1$. Using this value of τ and summing only up to $m - 1$ the previous expression becomes,

$$\sum_{r=0}^{m-1} \binom{m}{r} (-M_1)^{m-r} M_r + M_m = 0.$$

Or,

$$M_m = - \sum_{r=0}^{m-1} \binom{m}{r} (-1)^{m-r} (M_1)^{m-r} M_r.$$

From the hypotheses it is known that $M_0 = \int \phi(x) dx = 1$. Therefore, $M_\rho = M_1^\rho$ for $\rho = 0, 1$, and with this knowledge it is easy to show that $M_\rho = M_1^\rho$ for $\rho = 2$:

$$M_m = - \sum_{r=0}^{m-1} \binom{m}{r} (-1)^{m-r} (M_1)^{m-r} M_1^r,$$

which holds for $m = 1, 2$. Combine the powers of M_1 to get,

$$M_m = -M_1^m \sum_{r=0}^{m-1} \binom{m}{r} (-1)^{m-r}.$$

But, this is nothing more than,

$$M_m = -M_1^m [(1 - 1)^m - 1],$$

or simply,

$$M_m = M_1^m,$$

where $m = 1, 2$. The proof is complete, since higher powers of m can be found by induction.

Appendix B: Moments of the Scaling Function

In this appendix the moments of $\phi(x)$ will be calculated in closed form. Begin with the definition of the scaling function,

$$\phi(x) = \sum_k \phi(2x - k).$$

Next, calculating the m -th moment of $\phi(x)$ yields,

$$\int \phi(x)x^m = \sum_k h_k \int \phi(2x - k)x^m dx.$$

Change the variable of integration such that $y = 2x - k$ to get,

$$\begin{aligned} \int \phi(x)x^m &= \sum_k h_k \int \phi(y)(1/2)^m(y+k)^m 1/2 dy, \\ &= (1/2)^{m+1} \sum_k h_k \int \phi(y)(y+k)^m dy. \end{aligned}$$

Now, recall the binomial theorem to get,

$$\int \phi(x)x^m = (1/2)^{m+1} \sum_k h_k \int \phi(y) \sum_{l=0}^m \binom{m}{l} y^l k^{m-l} dy$$

Rewrite the moments of $\phi(x)$ as $M_l = \int x^l \phi(x) dx$ to get,

$$M_m = (1/2)^{m+1} \sum_{l=0}^m \binom{m}{l} \sum_k h_k k^{m-l} M_l.$$

Now let $\mu_l = \sum_k h_k k^l$ to get

$$M_m = (1/2)^{m+1} \sum_{l=0}^m \binom{m}{l} \mu_{m-l} M_l.$$

Now, M_m can be defined in terms of M_i for $i = 0, \dots, m-1$:

$$M_m(2^{m+1} - 2) = \sum_{l=0}^{m-1} \binom{m}{l} \mu_{m-l} M_l.$$

Note that the moments μ_l can be found by direct calculation given that the Daubechies filter coefficients, h_k , are already known.

The moments of $\phi(x)$ can now be used to find the mapping from the evenly-spaced samples of a function $f(x)$ to the scaling function coefficients. In section 3

i	M_i	μ_i	c_i
0	1	2	.3661
1	.63395	1.2679	.6340

Table 7: Scaling function moments for the Daubechies 4 wavelet.

i	M_i	μ_i	c_i
0	1	2	-.0235
1	1.0054	2.0108	1.0426
2	1.0108	2.0216	-.0199
3	.90736	.5078	.0009

Table 8: Scaling function moments for the Daubechies 8 wavelet.

this mapping was denoted in matrix form as the matrix C . The elements c_i which define the rows of this matrix have already been given for the wavelet D_6 . The comparable results for the D_4 and D_8 wavelets are given in the accompanying tables.



REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE December 1993	3. REPORT TYPE AND DATES COVERED Contractor Report		
4. TITLE AND SUBTITLE ON THE DAUBECHIES-BASED WAVELET DIFFERENTIATION MATRIX			5. FUNDING NUMBERS C NAS1-19480 WU 505-90-52-01	
6. AUTHOR(S) Leland Jameson				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Institute for Computer Applications in Science and Engineering Mail Stop 132C, NASA Langley Research Center Hampton, VA 23681-0001			8. PERFORMING ORGANIZATION REPORT NUMBER ICASE Report No. 93-95	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Langley Research Center Hampton, VA 23681-0001			10. SPONSORING/MONITORING AGENCY REPORT NUMBER NASA CR-191583 ICASE Report No. 93-95	
11. SUPPLEMENTARY NOTES Langley Technical Monitor: Michael F. Card Final Report Submitted to SIAM Journal on Scientific Computing				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified-Unlimited Subject Category 64			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The differentiation matrix for a Daubechies-based wavelet basis will be constructed and 'superconvergence' will be proven. That is, it will be proven that under the assumption of periodic boundary conditions that the differentiation matrix is accurate of order $2M$, even though the approximation subspace can represent exactly only polynomials up to degree $M - 1$, where M is the number of vanishing moments of the associated wavelet. It will be illustrated that Daubechies-based wavelet methods are equivalent to finite difference methods with grid refinement in regions of the domain where small-scale structure is present.				
14. SUBJECT TERMS differentiation matrix; superconvergence; wavelet approximation			15. NUMBER OF PAGES 53	
			16. PRICE CODE A04	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

National Aeronautics and
Space Administration
Langley Research Center
Mail Code 180
Hampton, VA 23681-0001

Official Business
Penalty for Private Use, \$300

BULK RATE
POSTAGE & FEES PAID
NASA
Permit No. G-27

